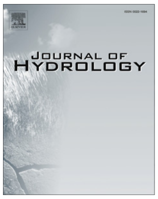
Journal of Hydrology 555 (2017) 518–534

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol



Research papers

Uncertainty-based simulation-optimization using Gaussian process emulation: Application to coastal groundwater management

Mohammad Mahdi Rajabi [a,*], Hamed Ketabchi [b]

[a] Civil and Environmental Engineering Faculty, Tarbiat Modares University, PO Box 14115-397, Tehran, Iran
[b] Department of Water Resources Engineering, Tarbiat Modares University, PO Box 14115-336, Tehran, Iran

ARTICLE INFO

Article history:
Received 30 March 2017
Received in revised form 15 October 2017
Accepted 19 October 2017
Available online 23 October 2017

This manuscript was handled by P. Kitanidis, Editor-in-Chief, with the assistance of Jian Luo, Associate Editor

Keywords:
Seawater intrusion
Uncertainty
Gaussian process emulation
Monte Carlo simulation
Continuous ant colony optimization
Persian Gulf

ABSTRACT

Combined simulation-optimization (S/O) schemes have long been recognized as a valuable tool in coastal groundwater management (CGM). However, previous applications have mostly relied on deterministic seawater intrusion (SWI) simulations. This is a questionable simplification, knowing that SWI models are inevitably prone to epistemic and aleatory uncertainty, and hence a management strategy obtained through S/O without consideration of uncertainty may result in significantly different real-world outcomes than expected. However, two key issues have hindered the use of uncertainty-based S/O schemes in CGM, which are addressed in this paper. The first issue is how to solve the computational challenges resulting from the need to perform massive numbers of simulations. The second issue is how the management problem is formulated in presence of uncertainty. We propose the use of Gaussian process (GP) emulation as a valuable tool in solving the computational challenges of uncertainty-based S/O in CGM. We apply GP emulation to the case study of Kish Island (located in the Persian Gulf) using an uncertainty-based S/O algorithm which relies on continuous ant colony optimization and Monte Carlo simulation. In doing so, we show that GP emulation can provide an acceptable level of accuracy, with no bias and low statistical dispersion, while tremendously reducing the computational time. Moreover, five new formulations for uncertainty-based S/O are presented based on concepts such as energy distances, prediction intervals and probabilities of SWI occurrence. We analyze the proposed formulations with respect to their resulting optimized solutions, the sensitivity of the solutions to the intended reliability levels, and the variations resulting from repeated optimization runs.

1. Introduction

Groundwater withdrawal in coastal aquifers needs to be managed based on optimal management strategies in order to prevent seawater intrusion (SWI) and guarantee the sustainability of groundwater use (Werner et al., 2013; Ataie-Ashtiani et al., 2014). In recent years, a significant number of studies have shown that these optimal management strategies can be derived from combined simulation-optimization (S/O) schemes (e.g. Kourakos and Mantoglou, 2009; Ataie-Ashtiani et al., 2014; Ketabchi and Ataie-Ashtiani, 2015b,c). In S/O schemes, SWI numerical models can be employed to analyze the effect of various management alternatives on the coastal aquifer system; and the optimization algorithm performs a systematic search for improved management alternatives based on SWI model outputs.

Previous applications of S/O in the context of coastal groundwater management (CGM) have mostly relied on deterministic SWI simulations. This simplifying assumption is certainly questionable, knowing that SWI model outputs are inevitably prone to both epistemic and aleatory uncertainty (Carrera et al., 2010; Rajabi and Ataie-Ashtiani, 2014). The inherent uncertainty in SWI models may cause significantly different real-world outcomes for a management strategy, compared to what is initially expected. So a management strategy that is considered optimal using a deterministic SWI model may become non-optimal when uncertainty in the model outputs is considered. This shortcoming has been recognized by several key review papers in the field of coastal aquifer S/O (e.g. Sreekanth and Datta, 2015; Ketabchi and Ataie-Ashtiani, 2015b,c; Ketabchi et al., 2016). However, a survey of literature shows that in practice, uncertainty-based S/O has rarely been considered in CGM (Sreekanth and Datta, 2014; Ketabchi and Ataie-Ashtiani, 2015b). The few previously published work on S/O under uncertainty in coastal groundwater applications are presented in Table 1. In this table a number of key features of previous work have been reviewed. As shown in Table 1, uncertainty-based

---

* Corresponding author.
E-mail addresses: mmrajabi@modares.ac.ir (M.M. Rajabi), h.ketabchi@modares.ac.ir (H. Ketabchi).

https://doi.org/10.1016/j.jhydrol.2017.10.041
0022-1694/

Nomenclature

3D	three dimensional
ACO	ant colony optimization
Alt.MP	alternative formulation of management problem
ANN	artificial neural networks
CACO	continuous ant colony optimization
CDF	cumulative distribution function
CGM	coastal groundwater management
CLD	centered L_2 discrepancy
CPU	central processing unit
CV	coefficient of variation
DoE	design of experiment
EA	evolutionary algorithm
ESE	enhanced stochastic evolutionary
F_1, F_2	failure events
FITC	fully independent training conditional
GMZ	groundwater management zone
GP	Gaussian process
GPr	genetic programming
max	maximum
MCMC	Markov chain Monte Carlo
MCS	Monte Carlo simulation
min	minimum
MP	management problem
OLHS	optimized Latin hypercube sampling
OR	optimization run
PC	personal computer
PCE	polynomial chaos expansion
PI	prediction intervals
QoI	quantities of interest
RBO	reliability-based optimization
RL	reliability level
S/O	simulation-optimization
SWI	seawater intrusion
TSR	time saving ratio
UP	uncertainty propagation
A_i	surface area of the *i*th GMZ
C_{SW}	salinity concentration in seawater
C_T	a threshold for salinity concentrations
$\bar{E}_i$	average groundwater extraction rate of the *i*th GMZ
E_i^{opt}	optimized extraction rate of the *i*th GMZ

E_{sum}	total annual groundwater extraction volume
$D(,)$	energy distance between two probability distributions
D_{sum}	sum of energy distances
D_{sum}^{allow}	allowable sum of energy distances
$F()$	probability distribution
F_j	CDF of salinity concentration in the *j*th observation well
F_j^Z	CDF of salinity concentration in the *j*th observation well in the zero-extraction scenario
$G()$	probability distribution
K_L	permeability of the lower geological layer
K_U	permeability of the upper geological layer
N_{GMZs}	total number of GMZs
N_{MC}	number of MCSs
$N_{MC,(C>C_T)}$	number of MCSs resulting in salinity concentrations above C_T
N_{no}	a number of successive generations in CACO
N_{Obs}	total number of observation wells
N_{tot}	total number of generations in CACO
P_F	probability of occurrence of a failure event
$R_{Net,i}$	net recharge in the *i*th GMZ
PI^L	lower bound of PI
PI^U	upper bound of PI
$Pr_{exd,j}^{C_T}$	the probability of exceedance of salinity concentrations in the *j*th observation well from a certain threshold
$Pr_{SWI,j}$	probability of SWI occurrence in the *j*th observation well
T_{GP}	time required to perform a set of calculations using the GP emulators
T_S	time required to perform a set of calculations using the numerical simulator
V	variance
X, X_i, X_j	model inputs
α_L	longitudinal dispersivity
α_T	transverse dispersivity
γ	design variables
δ	tolerable probability of failure
θ	uncertain inputs
μ	mean
σ	standard deviation

CGM studies have been limited to hypothetical examples (e.g. Dhar and Datta, 2009a,b; Sreekanth and Datta, 2011; Sreekanth et al., 2016) or simplified real-world cases (e.g. Sreekanth and Datta, 2014, Zekri et al., 2015). All the studies mentioned in Table 1 have used a particular evolutionary algorithm (EA), namely NSGA-II, for the optimizations. These studies have considered uncertainty in hydraulic conductivities, aquifer recharge and pumping rates.

Our survey shows that further investigations are necessary on several key aspects pertaining to S/O under uncertainty for CGM. More specifically, there are two important issues that require the most attention. The first issue is how to solve the computational challenges resulting from the need to performed massive numbers of numerical SWI simulations in S/O schemes that consider uncertainty. The second issue is how the management problem is formulated in presence of uncertainty. In the following subsections, these two issues and the potential strategies for addressing them are briefly described.

1.1. The computational challenge

The computational challenge of performing optimization under uncertainty arises from two key issues. First, estimation of the associated probabilistic objective function(s) or constraints generally requires the use of numerical methods such as Monte Carlo simulation (MCS) (Schueller et al., 2004), which involve large numbers of model simulations in order to reach the desired level of accuracy (Rajabi et al., 2015a). Second, the optimization procedure itself involves repeated evaluation of the objective function(s) and constraints in order to identify the optimal solution(s), and so, many repetitions of MCS are needed (Schueller and Jensen, 2008). Due to these reasons, it is computationally difficult to employ groundwater numerical models (which we hereafter refer to as the "simulators") in uncertainty-based S/O even in simple toy problems (He et al., 2010; Sreekanth, and Datta, 2014). Hence, in order to make simulation-optimization under uncertainty a practical tool for real-world groundwater management problems, it is necessary to find techniques that can substantially reduce the computational burden (Ketabchi and Ataie-Ashtiani, 2015b).

There are a number of potential strategies for reducing the computational cost of S/O under uncertainty, which include the use of: (1) efficient optimization techniques, (2) parallelization and grid computing, (3) efficient Monte Carlo methods, and (4) approximation techniques in the form of lower-fidelity or response surface surrogate modeling (Schueller and Jensen, 2008; Shan and Wang,